Amendments to the Claims:

This listing of claims will replace all prior versions, and listings, of claims in the application.

Listing of Claims:

1. (Currently Amended) A method of extracting new words automatically, said method comprising the steps of:

segmenting a cleaned corpus in a domain to form a segmented corpus;

splitting the segmented corpus to form sub strings, and counting the occurrences of each sub string[[s]] appearing in the corpus; and

filtering out false candidates to output new words, wherein the new words are words not contained in a base vocabulary;

wherein the segmenting and the splitting is not dependent upon word boundaries;

wherein new words are determined based upon the domain of the cleaned corpus;

wherein the step of splitting and counting is implemented using a GAST contained in a reduced memory space;

wherein a GAST is implemented by limiting length of character sub strings.

2. (Currently Amended) The method of extracting new words automatically according to **Claim 1**, wherein the step of segmenting comprises using punctuations, Arabic digits and alphabetic strings, or new word[[s]] patterns to split the cleaned corpus.

3. (Previously Presented) The method of extracting new words automatically according to **Claim 1**, wherein the step of segmenting comprises using common vocabulary to segment the cleaned corpus.

4. (Canceled)

5. (Canceled)

6. (Previously Presented) The method of extracting new words automatically according to **Claim 1**, wherein the step of filtering out false candidates comprises:

filtering out functional words;

filtering out those sub strings which almost always appear along with a longer sub string; and

filtering out those sub strings for which the occurrence is less than a predetermined threshold.

7. (Previously Presented) The method of extracting new words automatically according to **Claim 1**, wherein the step of segmenting the cleaned corpus comprises using pre-recognized functional words as segment boundary patterns.

8. (Previously Presented) The method of extracting new words automatically according to **Claim 3**, wherein the step of segmenting cleaned corpus comprises using pre-recognized functional words as segment boundary patterns.

9. (Currently Amended) The method of extracting new words automatically according to **Claim 3**, wherein the step of filtering out false words comprises:

filtering out functional words;

filtering out those sub strings which almost always appear along with a longer sub string[[s]]; and

filtering out those sub strings for which the occurrence is less than a predetermined threshold.

10. (Currently Amended) An automatic new word extraction system, comprising:

a segmentor which segments a cleaned corpus in a domain to form a segmented corpus;

a splitter which splits the segmented corpus to form sub strings, and which counts the number of the sub strings appearing in the corpus; and

a filter which filters out false candidates to output new words, wherein the new words are words not contained in a base vocabulary;

wherein the segmenting and the splitting is not dependent upon word boundaries;

wherein new words are determined based upon the domain of the cleaned corpus;

wherein the splitter builds a GAST contained in a reduced memory space;

wherein the GAST limits the length of character sub strings.

11. (Currently Amended) The automatic word extraction system according to **Claim 10**, wherein the segmentor uses punctuations, Arabic digits and alphabetic strings, or new word patterns to segment the cleaned corpus.

12. (Original) The automatic word extraction system according to **Claim 10**, wherein the segmentor uses common vocabulary to segment the cleaned corpus.

13. (Canceled)

14. (Canceled)

15. (Original) The automatic word extraction system according to **Claim 10**, wherein the filter filters out functional words; those sub strings which almost always appear along with longer sub strings; and those sub strings for which the occurrence is less than a predetermined threshold.

16. (Original) The automatic word extraction system according to **Claim 10**, wherein the segmentor uses pre-recognized functional words as segment boundary patterns.

17. (Original) The automatic word extraction system according to **Claim 12**, wherein the segmentor uses pre-recognized functional words as segment boundary patterns.

18. (Currently Amended) The automatic word extraction system according to **Claim 12**, wherein the filter filters out functional words; those sub strings which almost

always appear along with a longer sub string[[s]]; and those sub strings for which the occurrence is less than a predetermined threshold.

19. (Currently Amended) A program storage device readable by machine, tangibly embodying a program of instructions executable by the machine to perform method steps for extracting new words automatically, said method comprising the steps of:

segmenting a cleaned corpus in a domain to form a segmented corpus;

splitting the segmented corpus to form sub strings, and counting the occurrences of each sub string[[s]] appearing in the corpus; and

filtering out false candidates to output new words, wherein the new words are words not contained in a base vocabulary;

wherein the segmenting and the splitting is not dependent upon word boundaries;

wherein new words are determined based upon the domain of the cleaned corpus;

wherein the step of splitting and counting is implemented using a GAST contained in a reduced memory space;

wherein a GAST is implemented by limiting length of character sub strings.